HCEO WORKING PAPER SERIES

Analyzing Matching Patterns in Marriage: Theory and Application to Italian Data

Pierre-André Chiappori
Edoardo Ciscato
Carla Guerriero

Working Paper

2020-080

11/2020

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Analyzing matching patterns in marriage: theory and application to Italian data

Pierre-André Chiappori* Edoardo Ciscato† Carla Guerriero‡

October 26, 2020

Abstract

Social scientists have long been interested in marital homogamy and its relationship with inequality. Yet, measuring homogamy is not straightforward, particularly when we are interested in studying sorting on multiple traits. In this paper, we compare different statistical methods that have been used in the demographic, sociological and economic literature. We show that Separate Extreme Value (SEV) models not only generate a matching function with several desirable theoretical properties, but they are also particularly suited for the study of multidimensional sorting. We use small-scale survey data to study sorting among parents of children attending schools in Naples. Our findings show that homogamy is pervasive: not only men and women sort on age, education, height and physical characteristics, but they also look for partners that share similar health-related behavioral traits and risk attitude. We also show that marital patterns are well explained by a low number of dimensions, the most important being human capital. Moreover, children of parents with a high human capital endowment perform better at school, although they report lower levels of subjective well-being and perceived quality of relationship with their parents.

Keywords: homogamy, matching, intergenerational inequality.

*Columbia University, Department of Economics
†KU Leuven, Department of Economics
‡University of Naples Federico II, Department of Economics and Statistics

1 Introduction

Since the pioneering work of Becker (1973), a large number of studies have analyzed matching patterns in the marriage market. From a social sciences perspective, there are several reasons to study the choice of a partner, with a major motivation being the relationship between marital patterns and inequality. Assortative matching has a direct impact on inequality *within generations*: if individuals have a large propensity to marry their own likes, initial inequalities in individual endowments tend to be amplified at the household level (Fernández and Rogerson, 2001; Greenwood et al., 2003, 2014). More importantly, recent studies have emphasized the potential impact of assortative matching on social reproduction and *intergenerational* inequality. Individuals with a high level of human capital tend to match assortatively, and these households tend to invest heavily in their children's human capital (HC). Given the complementarities involved in the HC production function, such trends amplify initial inequalities for the next generation, generating what has sometimes be called an "inequality spiral".[1]

Numerous studies in sociology and demography have investigated marital patterns.[2] Although the economic and sociological literatures are addressing similar questions, they tend to use very distinct empirical approaches. Sociological and demographic analyses often rely on a direct, reduced-form estimation of the prevailing "matching function"; for this purpose, these contributions typically utilize either a "harmonic mean" or a log-linear approach. Economists tend to prefer more structural models: the recent development of the so-called Separable Extreme Value (from now on, SEV) models provides an interesting illustration.

When choosing appropriate tools to study marital patterns, an important issue is the *dimensionality* of the matching process. While social scientists are often interested in a small number of specific traits (e.g., income, education or HC), the real-life process of marital matching is obviously much more complex, and involves a host of other aspects: age, race, religion, but also tastes and preferences, cultural background, physical attractiveness, etc. From a methodological viewpoint, whether the empirical strategy adopted can account for this diversity, and more importantly whether it can disentangle the respective impacts of multiple traits - especially when the latter appear to be correlated - is an important aspect.

Another issue is related to the type of data that can be used for this type of analysis. Several

[1] See for instance Del Boca et al. (2014), Chiappori et al. (2017), Chiappori et al. (2017) and Chiappori et al. (2020).

[2] A non-exhaustive list includes Schoen (1981); Qian and Preston (1993); Blackwell and Lichter (2004); Schwartz and Mare (2005); Bouchet-Valat (2014); Schwartz and Han (2014); Gonalons-Pons and Schwartz (2017).

studies consider one "marriage market"; i.e., they analyze marital patterns within a given population, under the assumption that any two individuals of opposite gender could possibly match. Empirical estimation, in this context, faces difficult identification problems.[3] Alternatively, some contributions consider several "marriage markets". For instance, one can study the evolution of matching patterns over a long time period, and consider only marriages between individuals belonging to the same cohort.

The goal of the present paper is twofold. We first present the SEV approach, putting particular emphasis on its empirical implementation in a multidimensional context. We describe how a simple extension of the basic SEV model, borrowed from Dupuy and Galichon (2014), allows one to estimate the interactions between the various traits characterizing the spouses. We discuss how one can also consider more restricted models in which these numerous traits only matter through a small number of unknown "factors" (or "indices"). Such an assumption can be tested, and the corresponding factors can be empirically estimated.

We then apply our methodology to a rich Italian dataset, a survey of parents of children aged 6 to 19 attending schools in the Campania region around Naples. The dataset includes socio-demographic variables (age, education) and anthropometric characteristics (height and weight). It also collects information on health-related behavior such as smoking and sports activity, plus an incentivized question to investigate the relative healthiness of eating habits. Finally, the survey collects psychometric information on risk behavior, with a focus on health and recreational risk.

We show that marital patterns are characterized by homogamy not only with respect to age, education and anthropometric measures, but also with respect to a large number of behavioral traits related to health and risk attitude. We also show that sorting is based on a low number of dimensions; in other words, a model with one or few "indices of attractiveness" would constitute a good approximation of the marriage market. When we look into the different sorting dimensions, we show that the first mainly captures the market segmentation in different age (and possibly school) cohorts. However, the second dimension describes sorting on HC: not only educated women tend to marry educated men, but sorting also brings together individuals that are health-conscious (e.g., that smoke less and have a preference for healthy food). Finally, we briefly analyze the consequences of these assortative matching patterns in terms of children outcomes. We show that parents with a high level of HC have children that perform better at school, although these same children report lower levels of subjective well-being and a worse relationship with their mothers.

[3] If only one "market" is observed, SEV models are exactly identified under parametric assumptions. On the contrary, multi-markets models are typically overidentified, which may potentially lead to specification tests.

2 Analyzing Marital Patterns: Theory

2.1 The Basic Framework

We start by introducing some notations. We consider two populations, men and women, each defined by a set of characteristics X, Y; in practice, we shall assume that each woman (resp. man) is fully described by a vector x (resp. y), in the sense that two individuals with the same vector of characteristics are considered as perfect substitutes on the marriage market. In line with the previous discussion, these sets are typically *multidimensional*; moreover, some characteristics may be *unobservable* to the econometrician, thus reflecting unobserved heterogeneity among individuals. The distributions of characteristics (including their correlation patterns) within each population are described by two measures μ^X and μ^Y on each space.

The matching can be described in either of two ways. One can, on the one hand, define a mapping from X to Y, indicating for each woman x the man y to whom she is matched (or no one if she is single). An alternative, and paradoxically more convenient method is to define a measure μ on the product space $X \times Y$, where $\mu(x, y)$ denotes the probability that Mrs. x is matched with Mr. y; obviously, μ is constrained by the fact that its marginal must coincide with μ^X and μ^Y respectively. Note, in particular, that the second method allows for Mrs. x to be matched with positive probability to several mates - reflecting the fact that, in a typical data set, women with identical characteristics may well have observationally different partners (and conversely).

2.2 The matching function approach

Several approaches have been used in sociological and demographic studies of marital patterns. Schoen (1981) introduced the *harmonic mean* approach; more recently, Schwartz and Mare (2005) refer to a *log-linear* approach. The basic principles of these frameworks can be described as follows. Assume that individuals belong to a small number of "classes"; for instance, we are analyzing matching by education, the latter being categorized in S levels. Let $\mu_{IJ,t}$ denote the number of couples of cohort t in which she (he) belongs to class I (J); let $\bar{\mu}^X_{I,t}$ ($\bar{\mu}^Y_{J,t}$) denote the total number of women (men) with education I (J) in cohort t, and $\mu^X_{I,t}$ ($\mu^Y_{J,t}$) the total number of wives (husbands) with education I (J) in cohort t (where $t \in \{1, ..., T\}$); therefore the difference $\mu^X_{I\emptyset,t} = \bar{\mu}^X_{I,t} - \mu^X_{I,t}$ (resp. $\mu^Y_{\emptyset J,t} = \bar{\mu}^Y_{J,t} - \mu^Y_{J,t}$) is the total number of single women (men) with education I (J) in cohort t.

In the harmonic mean approach[4] one studies the relationship

$$\frac{1}{\mu_{IJ,t}} = \alpha_{IJ,t}\left(\frac{1}{\mu^X_{I,t}} + \frac{1}{\mu^Y_{J,t}}\right) \tag{1}$$

where the $\alpha_{IJ,t}$ are parameters to be empirically estimated; intuitively, $\alpha_{IJ,t}$ can be interpreted as the "force of attraction', in cohort t, between females in category I and males in category J.

The log-linear approach, alternatively, starts from the remark that, under the benchmark of random matching, one would get:

$$\mu_{IJ,t} = \mu^X_{I,t} \times \mu^Y_{J,t} \Rightarrow \ln \mu_{IJ,t} = \ln \mu^X_{I,t} + \ln \mu^Y_{J,t}$$

Assortative matching implies that the actual number of couples with both spouses belonging to the *same* class be larger than what would be expected under random matching. For that reason, a *homogamy log-linear model* adds homogamy dummy variables $\eta_{IJ,t},\ t = 1, ..., T$ that take the value 1 if $I = J$ and 0 otherwise. One can then regress $\ln \mu_{IJ,t}$ over the marginals $\left(\ln \mu^X_{I,t}, \ln \mu^Y_{J,t}\right)$ and the η_t:

$$\ln \mu_{IJ,t} = a \ln \mu^X_{I,t} + b \ln \mu^Y_{J,t} + \sum_t c_t \eta_{IJ,t} \tag{2}$$

Under the null of random matching, the c_t coefficients would all be nil; positive, significant coefficients therefore indicate the presence of (positive) assortative matching. One can moreover consider the evolution of the coefficients across cohorts; if they get larger, one can conclude that homogamy increases over time.

A *crossings log-linear model* introduces latent variables $\eta_{I,t}$ that represent the difficulty of "crossing" educational level I for cohort t. In practice, one can then define the variables $\theta_{IJ,t}$ by

$$\theta_{IJ,t} = \left\{ \begin{array}{c} \sum_{q=J}^{I-1} \eta_{q,t} \quad \text{if } I > J \\ \sum_{q=I}^{J-1} \eta_{q,t} \quad \text{if } I < J \\ 0 \quad \text{if } I = J \end{array} \right|$$

and regress the $\ln \mu_{IJ,t}$ over the marginals $\left(\ln \mu^X_{I,t}, \ln \mu^Y_{J,t}\right)$ and the $\theta_{IJ,t}$ to get an estimate of the (log) odds $\eta_{I,t}$ for $I = 1, ..., S$ and $t = 1, ..., T$. In other words, homogamy models summarize the evolution of homogamy over two cohorts by a single number (the difference in the corresponding ηs), whereas crossing models consider the various education levels independently; as such, they can account for more complex evolutions (e.g., homogamy increasing at the top of the education distribution but not at the bottom).

[4] See Qian and Preston (1993) for an application

Both harmonic mean and log-linear models are tractable and easy to estimate. However, they raise several problems. They are particularly adequate to study matching on one specific trait (e.g., education), but become much less convenient in a multidimensional context. Another issue relates to the adding-up constraints that are implicit in the underlying model. For any cohort t, it should be the case that

$$\mu^X_{I,t} = \sum_J \mu_{IJ,t} \text{ and } \mu^Y_{J,t} = \sum_I \mu_{IJ,t} \tag{3}$$

reflecting the fact that the sum, over all male education classes J, of the number $\mu_{IJ,t}$ of couples with a I-educated wife and a J-educated husband must equal the total number of I-educated wives (and conversely). However, such restrictions are difficult to impose on log-regressions like (1) or (2), and are typically ignored.[5] Finally, matching functions like (1) or (2) implicitly assume that the number of couples with a I-educated wife and a J-educated husband only depends on the total number of I-educated wives and J-educated husbands, not on the number of individuals in *other* education classes. Yet, basically all theoretical models of matching strongly suggest the existence of "spillovers", by which a change in the size of one class (say, more K-educated women, $K \neq I$) affects *all* marital patterns, including those involving women within education class I.[6] In particular, the use of such models to generate counterfactual simulations ("what would marital patterns be at $t = T$ if the tendency to homogamy was the same as at $t = 0$") is highly problematic.

2.3 The SEV approach

2.3.1 Principle: logistic regressions

From an empirical perspective, the SEV approach can primarily be described as a series of discrete choice models. With the same notations as before, the probability that, in cohort t, a woman i, belonging to class I, marries a husband j in J can be approximated by the empirical frequency $p^t_{IJ} = \mu_{IJ,t}/\bar{\mu}^X_{I,t}$, whereas the probability that she remains single is proxied by $p^t_{I\emptyset} = \mu^X_{I\emptyset,t}/\bar{\mu}^X_{I,t}$. These probabilities can be analyzed as stemming from a discrete choice model. Specifically, assume that there exists a set of additively separable, latent random variables $\left(\alpha^J_i, J = \emptyset, 1, ..., S\right)$ such that

[5] One possible solution, for the log-linear model, is to use, as right-hand side regressor, $\bar{\mu}^X_{I,t}$, the total number of I-educated *women*; instead of $\mu^X_{I,t}$. Then the number of I-educated single women, $\mu^X_{I\emptyset,t}$, would be such that

$$\ln \mu^X_{I\emptyset,t} = \ln\left(\mu_{-I,t} - \sum_J \mu_{IJ,t}\right) = \ln\left(\bar{\mu}^X_{I,t} - \sum_J \left(\mu^X_{I,t}\right)^{\alpha}\left(\mu^Y_{J,t}\right)^{\beta} \exp\left(\gamma\eta_t\right)\right)$$

However, why the dynamics of singlehood by education should take this intricate form (while the dynamics of marriage has the simple, log-additive form (2)) is not clear.

[6] Technically, the "matching function', which describes the relationship between the marginal distributions of male and female education and the resulting marital patterns, is typically such that the size of each cell depends on all marginals and not only on the size of the corresponding row and column, See for instance Chiappori et al. (2017, Chapter 4).

the utility for Mrs. i of choosing a husband with education J (resp. of remaining single) is $U^{IJ}+\alpha_i^J$ (resp. $U^{I\emptyset}+\alpha_i^\emptyset$) for some parameters U^{IJ}. Intuitively, U^{IJ} is the systematic utility generated, for a woman i in I, by marrying a husband in J, and the vector α_i describes i's idiosyncratic preferences for her husband's education. Then she marries a husband in J if and only if:

$$U^{IJ}+\alpha_i^J=\max_K\left(U^{IK}+\alpha_i^K\right),\quad K=(\emptyset,1,...,S)$$

implying that:

$$p_{IJ}^t=\Pr\left(\alpha_i^J-\alpha_i^K\geq U^{IK}-U^{IJ},\quad \forall K=(\emptyset,1,...,S)\right)$$

One can normalize $U^{I\emptyset}$ to be zero for all I; then the thresholds $U^{IJ}, J=1,...,K$ can be recovered by standard discrete regressions. For instance, assuming, as in Choo and Siow (2006) and Chiappori et al. (2017), that the αs are type I extreme value distributed (so that the differences $\alpha_i^J-\alpha_i^K$ are logistic), then the model is a standard multilogit one. Such a regression can be performed for each category of education. A similar analysis on the husband's side gives that Mr. j marries a wife with education I if and only if:

$$V^{IJ}+\beta_j^I=\max_K\left(V^{KJ}+\beta_j^K\right),\quad K=(\emptyset,1,...,S)$$

where the vector β_j describes j's idiosyncratic preferences for his wife's education. Again, assuming that the βs are extreme value distributed and under the normalization $V^{\emptyset J}=0\ \ \forall I$, the Vs can be recovered from standard mutilogistic regressions.

2.3.2 Structural interpretation

The set of (multi)logistic regressions is a natural empirical tool in this context. In addition, however, the SEV model offers a more structural interpretation. Specifically, assume that people match according to a frictionless matching game under transferable utility. That is (and omitting for the time being the cohort index t), any two individuals i and j generate, if they match, a surplus S_{ij} that can be shared between them; the TU assumption implies that individual excess[7] utilities *add up* to S_{ij}. The equilibrium concept is stability. A matching is stable when (i) no matched individual would rather be single, and (ii) no two individuals would both prefer being matched together rather than their current situation.

[7]Excess, here, refers to the additional gain received from marriage over and above the utility level the same person would reach as a single.

Assume, now, that the surplus takes the form:

$$S_{ij} = Z^{IJ} + \alpha_i^J + \beta_j^I$$

where the deterministic component Z^{IJ} only depends on the spouses' education classes, and the random shocks α and β have the same interpretation as before. The basic result, due to Choo and Siow (2006), states that the unique stable equilibrium of this game generates matching probabilities exactly equal to those generated by the series of logistic regressions just described if and only if

$$Z^{IJ} = U^{IJ} + V^{IJ}, \quad \forall I, J \in \{1, ..., S\}. \tag{4}$$

Conversely, the structural model can readily be identified by (i) running the series of regressions, and (ii) computing the structural components of the surplus using (4). Note that, unlike the harmonic mean or the log-linear frameworks, the SEV model mechanically satisfies the adding-up constraints discussed above; and spillover effects are paramount, since changing the size of any one class modifies all thresholds. Moreover, the model clearly distinguishes the respective impacts of the marginal distributions and of the structural tendencies towards assortative matching (as fully summarized by the Z^{IJ}s). This property is crucial for counterfactual simulations, as discussed below.

Finally, a useful property of logistic regressions, that extends to the SEV approach, is Independence from Irrelevant Alternatives (IIA), i.e., the fact that the *relative* probabilities of any set of possible choices do not depend on the presence of other ('irrelevant') alternatives. This implies, in particular, that estimating the probability of choosing a partner in education class I instead of K, *conditional on marriage*, will give the same conclusions as analyzing the *unconditional* choices (i.e., also taking singlehood as a possible choice).

2.3.3 Assortative matching

In the one-dimensional case, assortative matching is directly related to a specific property of the Z matrix, namely supermodularity. Take the case of matching on a (small) number of education classes. Positive assortative matching (from now on PAM) has the following, technical definition. Take four education levels I, I', J and J' such that $I > I'$ and $J > J'$, and consider all couples where the wife belongs to either I or I' and the husband belongs to either J or J'. Then PAM requires that there are more (I, J) and (I', J') couples (therefore less (I, J') and (I', J) ones) than would be expected under random matching. In particular, if $I = J$ and $I' = J'$, then PAM implies

more homogamous couples ((I, I) and (I', I')) and less heterogamous ((I, I') and (I', I)) ones than one would expect under random matching, which is a standard aspect of assortative matching.

It can be shown that matching patterns display positive assortativeness in that sense if and only if the matrix Z is *supermodular*, i.e. if and only if:

$$Z^{IJ} + Z^{I'J'} - Z^{IJ'} - Z^{I'J} \geq 0 \quad \text{for all } I < I' \text{ and } J < J' \tag{5}$$

This result is interesting since it provides a structural characterization for the important but sometimes vague notion of (positive) assortativeness. Note also that, in principle, assortativeness is a local property. For instance, it may be the case that matching is positive assortative at the top of the distribution but not at the bottom - in which case (5) is satisfied for upper education categories but not elsewhere.

2.3.4 Counterfactual simulations

Finally, a major advantage of the SEV approach is that it allows counterfactual simulations to be performed in a natural way. Consider two different periods - say, $t = 0$ and $t = T$. Between 0 and T, matching patterns have evolved, and these evolutions may have different causes. For one thing, the marginal distributions may have changed; for instance, the number of highly educated individuals has increased, and the variation may moreover be gender-specific (e.g., the number of educated women increased at a much faster pace). In itself, these changes would affect all matching patterns, as summarized by the μ_{IJ}. In addition, it may be the case that structural preferences for assortativeness have also evolved; for instance, marrying an educated spouse may be more important now than in the past, particularly for educated individuals. Indeed, several theoretical contributions have suggested that the huge increase in college premium that took place in the US over the last decades should result in higher benefits generated by endogamy, particularly at the top of the (education) distribution (Fernández et al., 2005; Chiappori et al., 2017). It is important, if only for policy purposes, to disentangle between the two effects. In particular, a natural, counterfactual question would be the following: "What would matching pattern at date T look like if preferences for endogamy were identical to those at date 0?".

In a SEV framework, the answer is straightforward. Indeed, under the assumptions made, structural preferences are fully described by the Z matrix at each date. The counterfactual simulation, therefore, simply require deriving the stable matching patterns that would emerge in a population where the distributions of characteristics by gender are those of date T (i.e., fully described by the

$\bar{\mu}^X_{I,T}$ and $\bar{\mu}^Y_{J,T}$), whereas the surplus are defined by the matrix $Z_0 = \left(Z_0^{IJ}\right)$. As it turns out, this derivation is quite simple, and exploits a standard property of matching model under transferable utility - namely, that stable matching maximize aggregate surplus (defined as the sum of pairwise surpluses S_{ij} for all matched pairs) over all possible matchings. Hence the stable matching can readily be derived as the solution to a linear maximization problem.[8]

2.4 Multidimensional SEV models: the quadratic case

The basic SEV model is non parametric, in the sense that it imposes no ex ante restriction on the Z coefficients, which are freely estimated from observed choices. As such, it can readily be extended to a multidimensional setting, where agents are defined by several characteristics: one can simply define classes as combinations of the various characteristics, and use the discrete choice approach described above. For instance, a class can be defined by an age range *and* an education level *and* a race, etc. However, such an approach often faces a dimensionality curse. Assuming that agents are defined by k characteristics, each of which may take l values, the total number of possible classes is k^l; even for relatively small values of k and l, this total may well exceed what can reasonably be estimated from a typical sample.[9]

In such a case, one may choose to introduce parametric restrictions on the structure of the model. A popular and simple one is to posit that the surplus is quadratic (Dupuy and Galichon, 2014). Thus, if the number of characteristics is m for women and n for men, so that a woman (a man) is fully described by a m-dimensional vector x (a n-dimensional vector y), the deterministic part of the surplus would be defined by a $m \times n$ matrix $A = ((a_{IJ}))$, called the *affinity matrix,* such that:

$$S(x, y) = x'Ay = \sum_{I,J} x_I y_J z_{IJ}$$

Note, in addition, that this formulation can accommodate discrete or continuous variables, and does not require the latter to be discretized, which can be useful for some applications.[10] The price to pay for that parametrization is that the estimation process can no longer rely on a series of independent logistic regressions; indeed, the quadratic structure would impose restrictions *across* the various multilogits. In practice, therefore, one can use a Maximum Likelihood Estimator that much resembles the one commonly used in the multinomial logit case, but that takes into account the multi-regression structure. The contribution of a couple to the likelihood function is given by

[8] See for instance Chiappori et al. (2020) for an example of such calculations.

[9] For instance, with $k = 5$ characteristics taking $l = 5$ possible values each, the number of classes is $5^5 = 3125$.

[10] Ordered discrete variables can be easily included in the framework detailed in this paper. Ciscato et al. (2020) discuss how to treat unordered discrete variables, such as race and ethnicity.

μ_{IJ}, the unconditional probability of observing a couple of type IJ in the data:

$$\hat{A} = \arg\max_{A} \frac{1}{N} \sum_{i} \mu_{I(i)J(i)}(A)$$

where $I(i)$ and $J(i)$ respectively denote the classes of a woman i and her partner, while N is the number of couples in our sample. In order to obtain the Maximum Likelihood Estimator $\hat{A}$, we need to calculate the predicted probability function μ for any Z considered when solving the optimization problem. The calculation of μ relies on fast and tractable numerical methods outlined in Appendix A.

2.5 Factor analysis in the SEV model

The quadratic specification offers another important advantage: we can rewrite the surplus S as the linear combination of independent "factors", each capturing a different dimension of assortativeness. This exercise is insightful for multiple reasons. First, it helps infer the number of dimensions of assortativeness: for instance, we can test the hypothesis that attractiveness is well summarized by a single index (or a small number of indices) subsuming numerous observable traits. Second, when we find that several dimensions of assortativeness matter, it helps quantify their relative importance. Third, it helps describe the role played by the observed variables x and y in each dimension of assortativeness. For instance, we may find that although men and women are each defined by a large number of characteristics, three (say) main indices are sufficient to capture a large fraction of the variance. These indices typically involve many (usually all) characteristics, each weighted by a specific coefficient; the analysis recovers these coefficients (therefore the "profiles" described by each index), and their respective contributions to total variance.

In practice, we need to perform a Singular Value Decomposition (SVD) of the affinity matrix Z.[11] We obtain

$$A = U' \Lambda V,$$

where Λ is a diagonal matrix whose nonincreasing elements $(\lambda_1, \ldots, \lambda_K)$, with $K = \min\{m, n\}$, tell us about the relative importance of each sorting dimension, while the columns of U and V are loading vectors that tell us about the nature of each dimension. In other words, we can define the

[11] Importantly, it is convenient to work with demeaned and rescaled data. More precisely, each observable characteristic x_k must be demeaned so that the sample mean $\hat{\mu}_k$ is zero; and all characteristics x must be rescaled so that the diagonal elements of the sample covariance $\hat{\Sigma}$ are one (i.e., $\hat{\sigma}_k = 1$ for every k). A similar transformation must be applied to y. We use this normalization in what follows.

indices of attractiveness $\tilde{x} = Ux$ and $\tilde{y} = Vy$ and rewrite the surplus as

$$x'Ay = \tilde{x}'\Lambda\tilde{y} = \sum_{k=1}^{K} \lambda_k \tilde{x}_k \tilde{y}_k$$

where each k term $\lambda_k \tilde{x}_k \tilde{y}_k$ represents the surplus contribution of an independent dimension of assortativeness.

In our empirical analysis, we can perform SVD on $\hat{A}$ to obtain estimates of U, V and the λs. In this way, we can discuss the relative importance and nature of the different dimensions of assortativeness. In this paper, we obtain confidence intervals for U, V and the λs by bootstrapping.

A natural question is how many relevant dimensions of assortativeness we observe, or how many λs are (significantly) positive. Dupuy and Galichon (2014) outline a method to answer this question and develop a test of joint significance of the estimated λs. In summary, the method consists in testing the rank of the estimated affinity matrix $\hat{A}$: the null hypothesis is a restriction on the rank of Z, i.e., $rank(A) = k$; if it is rejected, then the number of positive λs will be higher than k, which will lead us to conclude that the number of relevant dimensions of assortativeness is higher than k.[12]

3 Data

This study is part of a large research project, CHILDROLE, exploring the role of children as decision makers within the family. Data were collected from February to April 2019 in five schools located within the province of Naples. Naples is Italy's third-largest city and its province is one of the most densely populated in Europe. Naples and the surrounding towns are marked by wide income and cultural differences and offer a good setting for collecting representative sample data on Italian households. Families were recruited through schools that agreed to participate, classes being selected randomly to take part in the study. Five schools (three elementary, one middle school and one high school) agreed to take part. The schools, all public, are located in different districts of the city and nearby towns, with a good socioeconomic mix. The parent survey was by means of face-to-face interviews and pencil-and-paper questionnaires (booklets with numbered pages and large font text). To avoid reciprocal influence, fathers and mothers were asked to complete the questionnaire in separate rooms. The data contain detailed information about the household and its individual members.

[12] Another way of proceeding is to impose a restriction on the rank of A beforehand, as suggested by Dupuy et al. (2019). The interest of this methodology is to allow for A to be sparse and limit the number of parameters to estimate when the sample size is relatively small with respect to the number of variables.

3.1 Variables

Respondents were asked to report their date of birth (missing answers supplied from school registry of date and place of birth of both parents). Educational attainment corresponds to the self-reported respondent's highest level of education achieved: 1) Primary; 2) Intermediate, 3) High School, or 4) University. Self-reported height and weight are used to measure Body Mass Index (BMI).

Respondents were also asked about their health-related risk behaviors, measured by three variables: smoking, physical exercise, and propensity for healthy diet. Smoking and physical exercise are measured through multiple choice questions with three possible answers: "never" (coded as 0), "seldom" (1) and "often" (2). Propensity for healthy diet was assessed through an incentivized question: each parent had to choose a snack to consume after completing the questionnaire.[13] At the end of the interview a generic question investigated respondents' concern for their own health: "Do you worry about your own health?"

Next, information was collected on respondents' measures of domain-specific risk-taking behavior. Weber et al. (2002) originally developed and tested an individual measure of risk taking in judgment and decision-making in different conventional domains: financial, ethical, health/safety, social, and recreational (Weber et al., 2002; Blais and Weber, 2006). We used only the questions related to the health and safety and the recreational domains. The questions were modified with the help of a child psychologist, in order to explore how risk measures vary among all the household members (including children and youths). After extensive piloting, additional questions applicable to both parents and children were included in the final questionnaire (see Appendix B); for further details on the questions, see Guerriero et al. (2018).

3.2 Summary Statistics

Table 1: Frequency of children by parents' survey participation

	Father participates		
Mother participates	**0**	**1**	**Total**
0	326	8	334
1	48	276	324
Total	374	284	658

Notes: participation is coded as the parent reporting at least some basic information.

Table 1 reports the number of households participating in the study: 632 children were con-

[13] Respondents were shown three different types of snack before the interview and were asked to pick one of them to collect immediately after the interview. The choices were: a banana, a Parmesan bar, and a chocolate muffin. According to the Center for Disease Control, the three correspond to different degrees of healthiness: respectively, very healthy, healthy and unhealthy.

tacted; for 332 of them, at least one parent participated to the survey; for 276 of them both parents participated to the survey. The latter constitute the core of our sample, since information on both parents are necessary for our empirical analysis. Table 2 reports summary statistics on parents' education and age. We have complete information on age, educational attainment, height and BMI for 259 out of 276 participant couples. Table 3 reports summary statistics on parents' health and recreational risk behavior: for 196 out of 276 participant couples, both parents completed the behavioral questionnaire. In table 2, we see that men are on average 43 year old and women 40 year old. Women are also slightly better educated, a result is in line with Italian national statistics (Eurostat, 2016). The average man in our sample is overweight according to WHO standards (BMI greater than 25.0), while women are on average in the healthy range (BMI between 18.5 and 24.9). Table 3 shows summary statistics on risk behavior. Women smoke less, but are also more inactive; preferences for healthy snacks are on average similar across genders, and so are concerns about own health. Gender differences are more pronounced for specific health and recreational risk behaviors: women are more likely to use sunscreen at the beach and dislike fast driving and extreme sports. On the other hand, women are less likely to wear a motorbike helmet than men.

Table 2: Summary statistics - parents

	count	mean	sd	p10	p90
Mother's education	323	2.9	0.8	2.0	4.0
Mother's age	322	39.7	6.8	31.0	49.0
Mother's height (cm)	314	163.6	5.6	158.0	170.0
Mother's BMI	308	24.3	3.9	20.4	29.3
Father's education	281	2.7	0.8	2.0	4.0
Father's age	283	43.1	7.3	35.0	53.0
Father's height (cm)	279	175.3	6.6	168.0	183.0
Father's BMI	274	26.8	3.4	23.0	31.1

Notes: education is coded as a four-category variable. Height and BMI trimmed.

4 Results

4.1 Reduced affinity matrix

We start with a small-size version of our model, in which we only consider matching patterns on four socio-demographic characteristics: age, education, height and BMI. The corresponding matrix is given in Table 4. A striking property of the matrix is the high level of assortative matching it reveals: all diagonal coefficients are positive and significant, implying that for all characteristics, homogamy increases the surplus generated by the match. Not surprisingly, the largest and most significant association relates to age - a feature that essentially reflects the presence of several

Table 3: Summary statistics - parents

	mean	sd	count
Mother smokes	0.6	0.8	310
Mother does sports	0.8	0.8	310
Mother likes healthy snacks	0.8	0.8	320
Mother puts sunscreen	1.4	0.7	309
Mother washes hands	1.8	0.4	310
Mother worries about her own health	1.5	0.6	311
Mother wears helmet	1.6	0.7	285
Mother would do a safari	0.8	0.8	308
Mother hates speed	1.3	0.7	295
Mother likes usual vacation	1.1	0.7	307
Mother likes extreme sports	0.2	0.5	311
Mother crosses carefully	1.9	0.3	311
Father smokes	0.8	0.9	269
Father does sports	0.9	0.7	267
Father likes healthy snacks	0.8	0.8	277
Father puts sunscreen	1.0	0.8	270
Father washes hands	1.8	0.5	266
Father worries about his own health	1.5	0.6	269
Father wears helmet	1.7	0.6	257
Father would do a safari	0.8	0.8	267
Father hates speed	0.7	0.7	260
Father likes usual vacation	1.1	0.7	265
Father likes extreme sports	0.4	0.6	269
Father crosses carefully	1.8	0.5	268

Notes: possible answers are (0) never, (1) sometimes, (2) often. The only exception is the question about healthy snacks, for which possible answers are (0) chocolate muffin, (1) Parmesan bar, (2) banana.

Table 4: Estimated affinity matrix (sample 1)

Husband \ Wife	Educ.	Age	Height	BMI
Educ.	**0.74**	**0.62**	0.14	**-0.24**
	(0.11)	(0.17)	(0.09)	(0.10)
Age	0.24	**3.30**	0.25	0.07
	(0.16)	(0.33)	(0.13)	(0.13)
Height	**0.25**	**0.37**	**0.16**	-0.07
	(0.09)	(0.14)	(0.07)	(0.07)
BMI	0.10	0.20	0.01	**0.28**
	(0.08)	(0.13)	(0.07)	(0.07)

Notes: 259 couples. Standard errors in parentheses: the estimator $\hat{A}$ is asymptotically normally. Bold-faced values are significant at the 5% level.

cohorts among parents. Very significant as well is homogamy on education (which was expected), but also on BMI and height. Additional patterns emerge; for instance, more educated men tend to have older and thinner wives (while the opposite is not significant); and more educated wives tend to have taller husbands.

Table 5: Saliency analysis: men's loading matrix (sample 1)

	Index 1	Index 2	Index 3	Index 4
Educ.	**0.21**	**0.92**	0.03	**-0.32**
Age	**0.97**	**-0.23**	**-0.07**	-0.05
Height	**0.12**	**0.30**	0.01	**0.95**
BMI	**0.06**	-0.05	**1.00**	-0.01
Index share	0.75	0.17	0.06	0.02

Notes: the table reports men's singular vectors V and singular values $diag(\Lambda)$ from the singular value decomposition of $\hat{A} = U\Lambda V'$. Bold-faced values are significant at the 5% level; confidence intervals are obtained with 500 bootstrap replications (Milan and Whittaker, 1995). In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

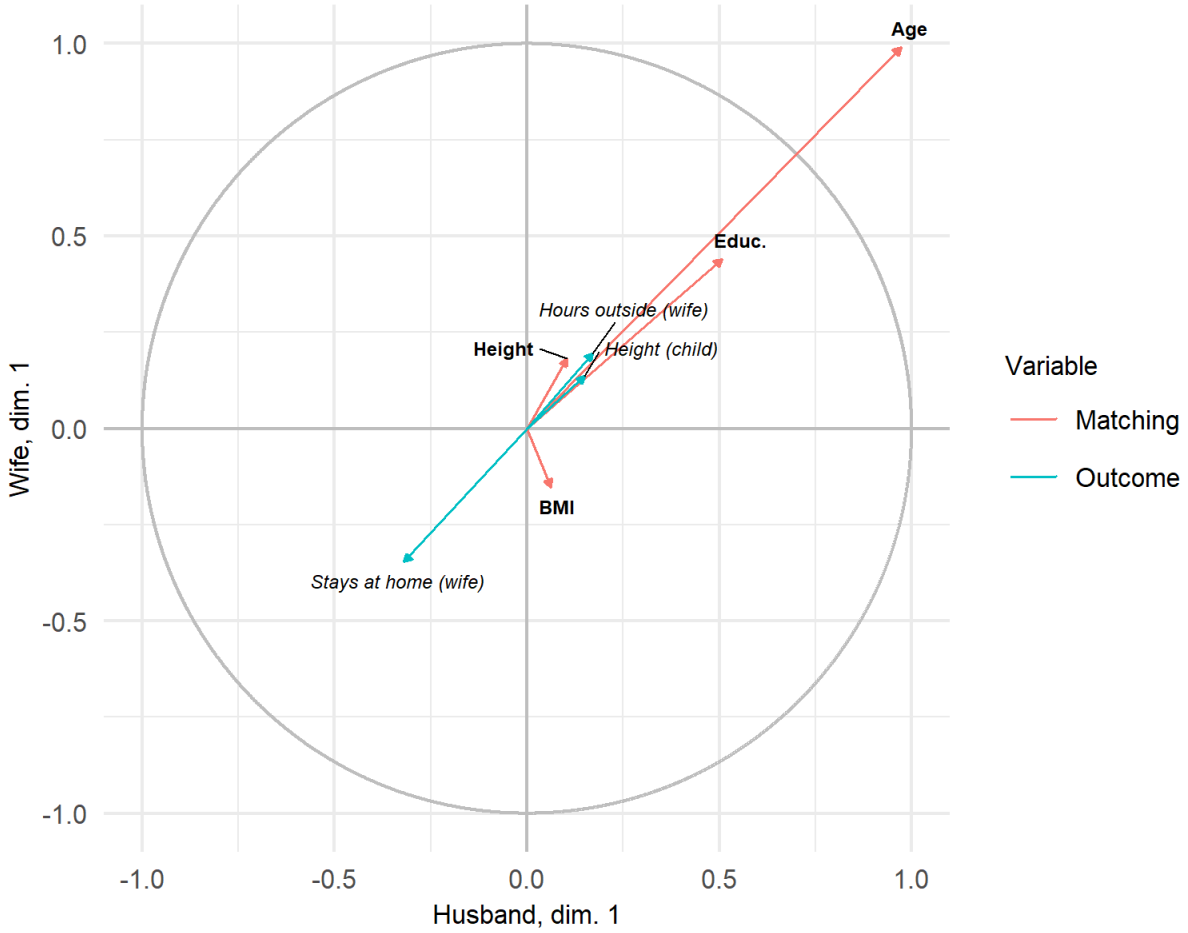
Table 6: Saliency analysis: women's loading matrix (sample 1)

	Index 1	Index 2	Index 3	Index 4
Educ.	**0.12**	**0.91**	**0.36**	-0.14
Age	**0.99**	**-0.12**	**-0.05**	-0.07
Height	**0.08**	**0.15**	-0.01	0.99
BMI	0.01	**-0.36**	**0.93**	0.07
Index share	0.75	0.17	0.06	0.02

Notes: the table reports women's singular vectors U and singular values $diag(\Lambda)$ from the singular value decomposition of $\hat{A} = U\Lambda V'$. Bold-faced values are significant at the 5% level; confidence intervals are obtained with 500 bootstrap replications (Milan and Whittaker, 1995). In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

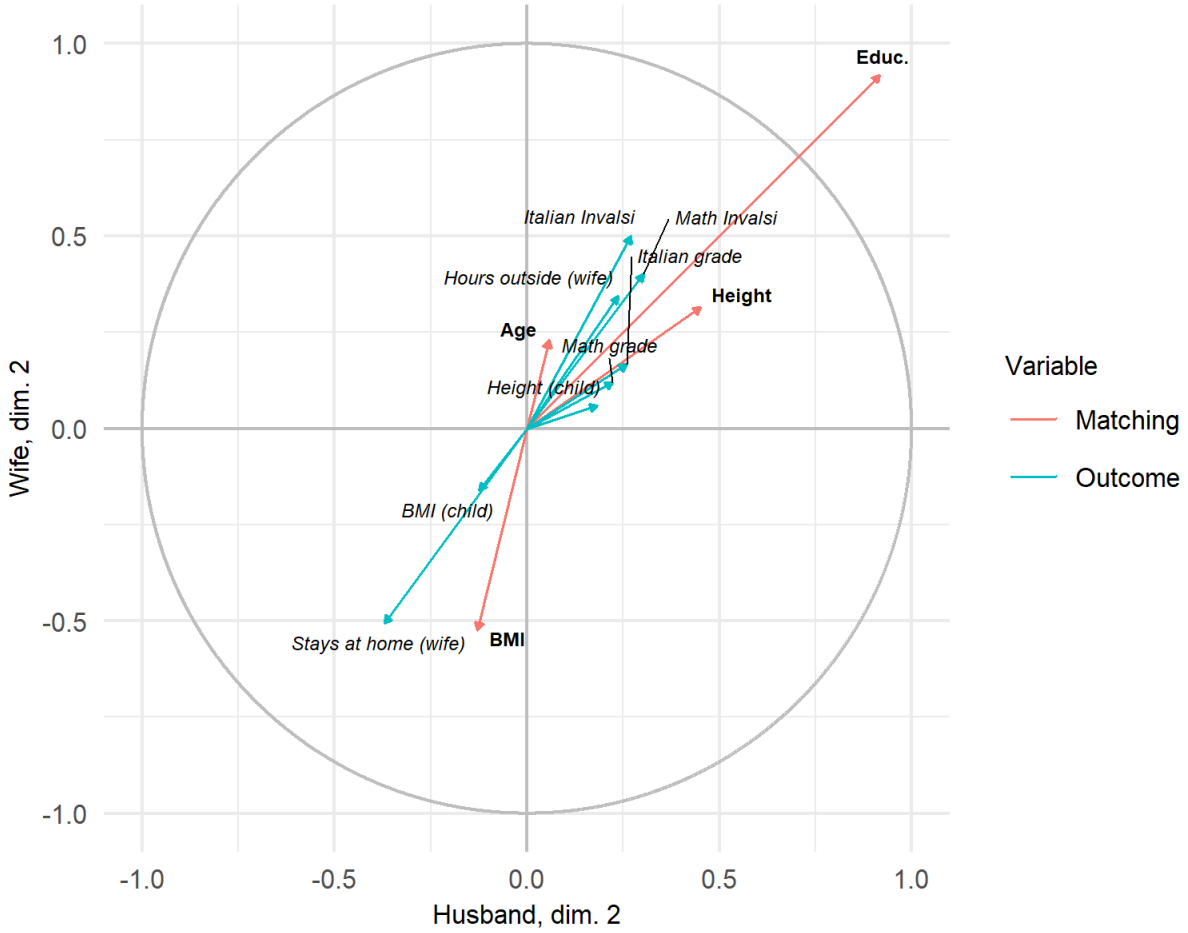
The factor decomposition is also interesting; it is given in Tables 5 and 6. The first factor explains three quarters of the total explained variance for both men and women, and mostly reflects age differences. In other words, parents in our sample belong to different "cohorts" (defined by year of birth), and people tend to marry a spouse from a cohort that is close to their own. The second factor (representing 17% of total variance for both genders) is more interesting. It singles out individuals who are more educated, as well as taller and (at least for women) thinner. In other words, matching patterns, while primarily driven by age, also capture a mix of education and physical appearance, possibly reflecting various dimensions of social status. Figures 1 and 2 provide a graphical visualization of the first two matching dimensions and help us with the interpretation: we will get back to them in the next sections.

Figure 1: Correlation between index 1, matching variables and outcomes (sample 1)



Notes: we plot correlation rates of both matching variables and outcome variables with the husband's first matching index (x-axis) and the wife's (y-axis). The matching variables include all the parents' background variables used in the main estimation. The outcomes include additional variables that are excluded from the main estimation: e.g., the number of children, the child's grades, the wife's labor supply. In order to improve the readability of the graph, we only plot those variables whose correlation rate is significantly different from zero at the 2% level.

Figure 2: Correlation between index 2, matching variables and outcomes (sample 1)



Notes: we plot correlation rates of both matching variables and outcome variables with the husband's second matching index (x-axis) and the wife's (y-axis). See notes to figure 1.

4.2 Global affinity matrix

We now consider the global (16×16) affinity matrix. A first and very striking feature is the high level of *homogamy* that prevails within the population. Each of the 16 diagonal coefficients in Table 7 is positive, indicating positive assortativeness along that specific dimension; the probability of getting such a pattern under random matching would be less than .01%. Moreover, each but one is statistically significant at 5%, and most are actually significant at 1%. This is all the more remarkable given the relatively small sample size (less than 200 couples).[14]

Several aspects revealed in the reduced matrix are still visible here - for instance, a positive (and significant) interaction between his education and her age, or between her education and his height. Other are less expected. Wives of older and more educated men are less likely to favor

[14]If the spouses become more similar in terms of health-related behavior and risk attitude *after* the marriage, our estimates of the diagonal coefficients would be upward biased. If we had information on the length of the relationship, we would be able to run a robustness check with newlyweds only. However, there is already evidence in the literature that this kind of traits do not change significantly after marriage: Dohmen et al. (2012) find that correlation patterns between spouses' risk attitude and trust do not change with the length of the relationship.

vacations at an unusual place. Husbands of more educated women are less likely to do a lot of sport unless they are extreme; they are also more likely to eat healthy food.

Table 7: Estimated affinity matrix (sample 2)

Husband \ Wife	Educ.	Age	Height	BMI	Smokes	Likes sports	Likes healthy snacks	Wears sunscreen	Wash hands	Worries about health	Wears helmet	Likes safari	Fear speed	Likes usual holidays	Extreme sports	Careful when crossing
Educ.	**0.99**	**0.87**	0.15	-0.30	-0.04	0.11	0.26	-0.10	0.07	0.12	-0.19	-0.05	0.28	**-0.40**	0.19	-0.32
	(0.19)	(0.27)	(0.14)	(0.15)	(0.16)	(0.14)	(0.16)	(0.18)	(0.16)	(0.15)	(0.18)	(0.15)	(0.16)	(0.15)	(0.14)	(0.17)
Age	0.21	**4.37**	0.37	0.31	0.17	0.04	0.05	**0.57**	0.02	0.02	0.31	-0.10	-0.13	**-0.43**	0.28	-0.16
	(0.24)	(0.49)	(0.20)	(0.21)	(0.22)	(0.20)	(0.22)	(0.26)	(0.21)	(0.20)	(0.24)	(0.22)	(0.22)	(0.21)	(0.20)	(0.23)
Height	**0.31**	0.24	**0.30**	-0.08	0.11	0.01	0.00	0.06	-0.04	-0.16	-0.12	0.13	0.03	0.13	0.01	0.07
	(0.13)	(0.20)	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.13)	(0.13)	(0.11)	(0.13)	(0.11)	(0.12)	(0.11)	(0.10)	(0.13)
BMI	0.03	0.06	0.05	**0.57**	-0.19	-0.08	-0.07	0.25	0.18	-0.00	0.15	-0.07	-0.04	-0.10	-0.08	-0.24
	(0.13)	(0.21)	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.13)	(0.15)	(0.11)	(0.13)	(0.12)	(0.12)	(0.11)	(0.11)	(0.14)
Smokes	-0.13	0.05	-0.03	0.05	**0.51**	-0.03	-0.16	0.08	-0.05	0.03	-0.01	0.13	-0.16	-0.02	0.01	0.05
	(0.14)	(0.21)	(0.11)	(0.11)	(0.12)	(0.11)	(0.12)	(0.13)	(0.13)	(0.11)	(0.13)	(0.11)	(0.12)	(0.11)	(0.11)	(0.14)
Likes sports	**-0.35**	0.24	0.16	-0.02	-0.18	**0.29**	**-0.33**	**-0.38**	-0.24	0.08	0.22	0.17	-0.06	0.19	0.03	**0.38**
	(0.15)	(0.22)	(0.12)	(0.12)	(0.13)	(0.12)	(0.13)	(0.14)	(0.14)	(0.12)	(0.15)	(0.13)	(0.13)	(0.12)	(0.12)	(0.16)
Likes healthy snacks	**0.35**	-0.09	-0.03	0.03	0.13	-0.00	**0.71**	0.27	0.07	-0.10	-0.02	0.02	0.05	0.18	-0.22	-0.08
	(0.15)	(0.22)	(0.12)	(0.13)	(0.13)	(0.12)	(0.13)	(0.15)	(0.13)	(0.12)	(0.14)	(0.13)	(0.13)	(0.12)	(0.13)	(0.15)
Wears sunscreen	0.26	-0.30	-0.02	0.15	0.05	-0.17	**0.45**	**0.51**	-0.07	-0.04	0.07	-0.17	-0.24	0.12	0.14	**-0.35**
	(0.15)	(0.22)	(0.12)	(0.12)	(0.12)	(0.12)	(0.14)	(0.14)	(0.14)	(0.12)	(0.14)	(0.13)	(0.13)	(0.12)	(0.12)	(0.16)
Wash hands	**-0.51**	0.19	-0.04	0.07	-0.24	-0.09	-0.08	-0.14	**0.60**	0.01	0.11	-0.03	-0.15	0.03	-0.08	0.16
	(0.18)	(0.23)	(0.12)	(0.14)	(0.15)	(0.13)	(0.14)	(0.16)	(0.14)	(0.13)	(0.14)	(0.13)	(0.14)	(0.13)	(0.11)	(0.13)
Worries about health	-0.18	0.08	0.04	-0.04	-0.20	0.08	-0.11	-0.15	-0.32	**0.35**	0.19	-0.07	-0.06	-0.14	-0.15	0.11
	(0.15)	(0.23)	(0.12)	(0.13)	(0.13)	(0.12)	(0.13)	(0.14)	(0.17)	(0.12)	(0.14)	(0.13)	(0.13)	(0.12)	(0.12)	(0.16)
Wears helmet	0.20	-0.04	-0.00	-0.15	-0.20	0.07	-0.11	-0.13	**0.34**	-0.18	**0.68**	0.13	**-0.44**	0.18	-0.13	-0.22
	(0.18)	(0.26)	(0.15)	(0.13)	(0.15)	(0.15)	(0.15)	(0.16)	(0.14)	(0.14)	(0.15)	(0.15)	(0.16)	(0.15)	(0.14)	(0.16)
Likes safari	-0.09	-0.02	-0.05	-0.17	-0.14	0.08	-0.11	**-0.33**	0.12	0.09	-0.06	**0.58**	0.11	-0.05	0.06	0.05
	(0.14)	(0.22)	(0.11)	(0.13)	(0.13)	(0.11)	(0.12)	(0.14)	(0.13)	(0.11)	(0.14)	(0.12)	(0.12)	(0.11)	(0.11)	(0.15)
Fear speed	-0.00	0.11	0.03	0.10	**0.34**	0.09	0.14	0.17	-0.15	0.03	**-0.40**	-0.03	**0.53**	0.01	0.06	-0.13
	(0.14)	(0.21)	(0.11)	(0.11)	(0.12)	(0.11)	(0.12)	(0.13)	(0.12)	(0.11)	(0.13)	(0.11)	(0.13)	(0.11)	(0.11)	(0.14)
Likes usual holidays	0.11	0.10	-0.19	0.11	**-0.24**	-0.02	0.17	-0.22	-0.17	0.02	0.05	0.21	0.20	**0.53**	0.07	**-0.39**
	(0.14)	(0.21)	(0.11)	(0.11)	(0.12)	(0.11)	(0.13)	(0.13)	(0.13)	(0.11)	(0.13)	(0.12)	(0.12)	(0.11)	(0.11)	(0.16)
Extreme sports	**0.28**	0.19	0.08	-0.01	0.16	0.01	0.19	0.16	0.22	-0.01	-0.14	-0.11	0.19	0.00	0.02	-0.20
	(0.13)	(0.20)	(0.11)	(0.12)	(0.12)	(0.11)	(0.12)	(0.13)	(0.14)	(0.11)	(0.13)	(0.11)	(0.12)	(0.11)	(0.10)	(0.14)
Careful when crossing	-0.33	-0.10	0.07	-0.04	0.26	-0.06	0.07	0.14	-0.05	0.14	**-0.42**	-0.17	0.24	-0.12	0.01	**0.47**
	(0.17)	(0.24)	(0.14)	(0.13)	(0.16)	(0.14)	(0.15)	(0.16)	(0.14)	(0.13)	(0.16)	(0.14)	(0.14)	(0.14)	(0.12)	(0.13)

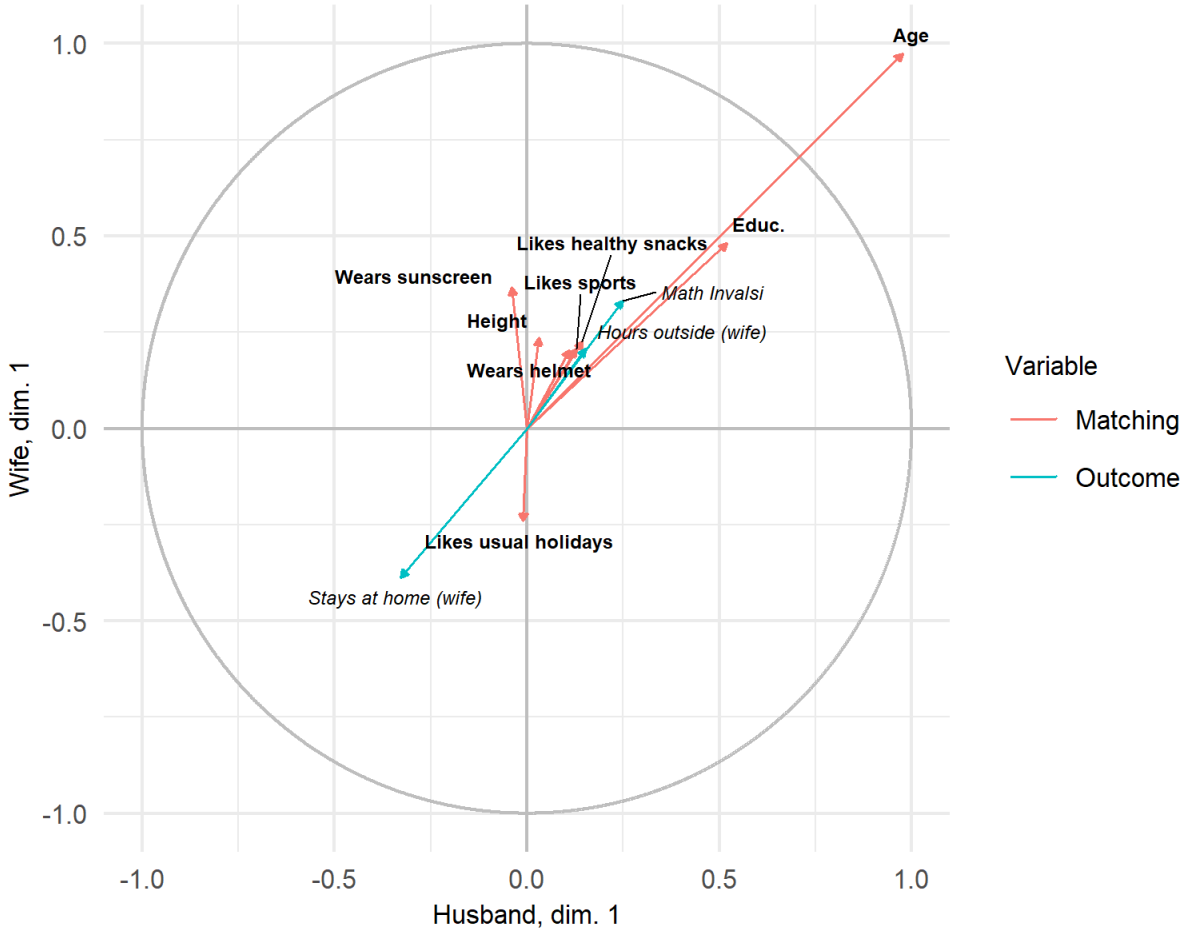
Notes: 196 couples. Standard errors in parentheses: the estimator $\hat{A}$ is asymptotically normally. Bold-faced values are significant at the 5% level.

More importantly, the factor decomposition enriches the conclusions drawn from the reduced matrix. In Figures 3, 4 and 5, we plot the correlation between men's and women's first three matching factors and their observable traits; factor loadings are reported in Tables 15 and 16 in Appendix C. The first factor - which, by itself, explains about a third of total explained variance - essentially recaptures the cohort pattern observed on the reduced matrix: in the first columns of Tables 15 and 16, we can see that age plays a dominant role in the first sorting dimension.

The second factor singles out differences in education: individuals with a high Index 2 appear to be more educated (as before), but also more health-conscious (they tend to eat healthy food and use sunscreen) and less likely to smoke or experience health problems. Along the second dimension of matching, some characteristics are gender-specific: women with a high Index 2 are thinner but also older; men with a high Index 2, are less likely to do sports and wash their hands before eating.

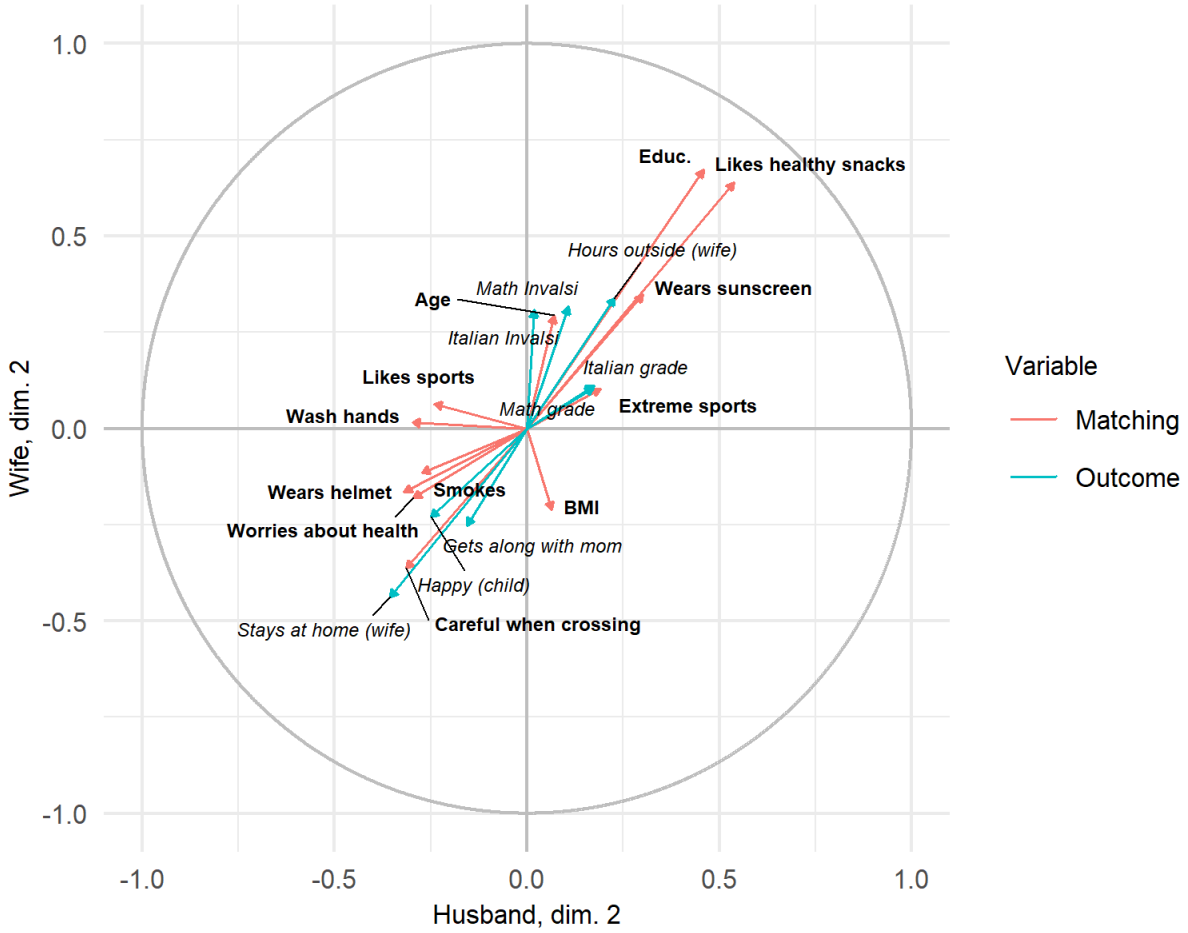
Finally, the third factor emphasizes traits that may relate to risk aversion. According to this third dimension, spouses tend to share a common attitude toward smoking, use of sunscreen and motorcycle helmet and hand washing; they also have a similar attraction (or dislike) for high speed and tend to pay more (or less) attention when crossing the street. Together, these three factors explain more than half the total (explained) variance - i.e., more than the remaining 13 factors combined.

Figure 3: Correlation between index 1, matching variables and outcomes (sample 2)



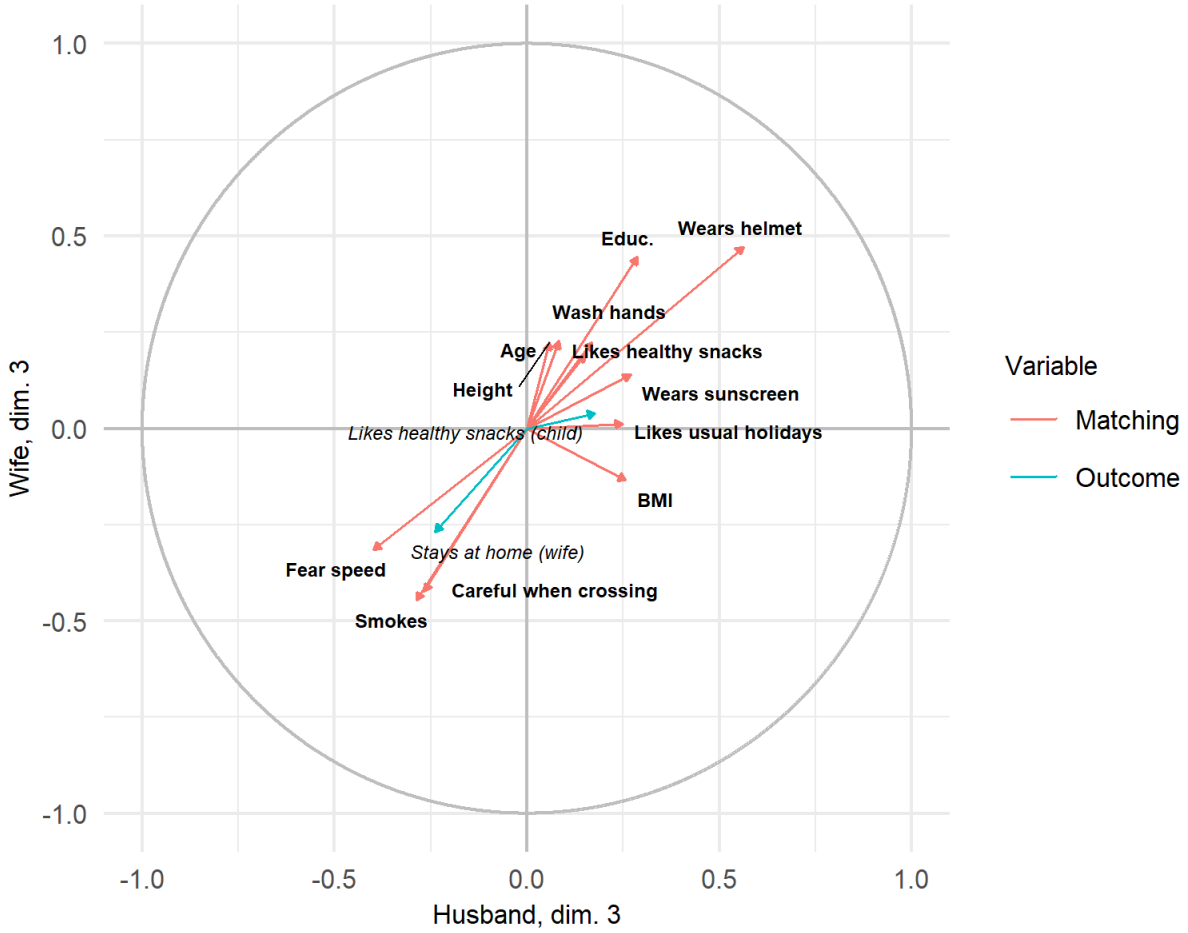
Notes: we plot correlation rates of both matching variables and outcome variables with the husband's first matching index (x-axis) and the wife's (y-axis). See notes to figure 1.

Figure 4: Correlation between index 2, matching variables and outcomes (sample 2)



Notes: we plot correlation rates of both matching variables and outcome variables with the husband's second matching index (x-axis) and the wife's (y-axis). See notes to figure 1.

Figure 5: Correlation between index 3, matching variables and outcomes (sample 2)



Notes: we plot correlation rates of both matching variables and outcome variables with the husband's third matching index (x-axis) and the wife's (y-axis). See notes to figure 1.

4.3 Impact on children

An interesting aspect of our data is that they include what could be considered as "outcome" variables - i.e., indicators reflecting choices made within the household and their consequences. A standard example is labor supply behavior: while most married men are active on the labor market, women may or may not participate, and these decisions appear to be related to matching patterns. Of particular interest is the impact of matching on children: here we measure both objective outcomes, such as grades and academic performance, and other subjective indicators, such as the child's subjective well-being and the perceived quality of the relationship with their parents. Note that the latter variables exhibit a clear age pattern, older children typically exhibiting less satisfaction and more difficult parental connections. For that reason, we often "clean up" these variable from their age-related component.[15] A list of all outcome variables is available in Appendix

[15] In practice, we regress the variable under consideration on the child's age, and we use the residual of this regression as our new indicator.

B.

Figure 1 is based on the reduced affinity matrix; it plots the composition of the wife's and husband's first index together with some output variables. In order to improve the readability of the graph, we only keep variables whose correlation with the corresponding factor is significant at the 2% level. Parents with high index 1 are older, which correlates with children reporting lower levels of subjective well being. This correlation, however, appears to be spurious, and mostly reflecting the child's age; once controlling for age (in the way explained just above), the correlation disappears. In other words, the cohort component of the matching patterns does not seem to be strongly correlated with any output, with the notable exception of women being more likely to work outside home - although even this correlation is likely to partly reflect the (older) age of the children.

Things are however quite different with the second factor, which is mostly driven by parents' education. In figure 2, we see that high levels of index 2 are strongly correlated with better educational outcomes, as measured by children's grades: the latter include both nationally standardized INVALSI test results (math and Italian) and class grades (always math and Italian, standardized at class level). Women with a high index 2 are also more likely to work outside home and less likely to be overweight, while both parents appear to be taller: children of parents with a high index 2 are also less likely to be overweight and taller than their peers. Finally, the third index mainly captures variations in parents' BMI; our results suggest that parents with a high index 3, as well as their children, are more likely to be overweight (see Figure 6 in Appendix D).

These conclusions are enriched when considering the global affinity matrix (Figures 3, 4 and 5). As before, the first index is only correlated with women's participation to the labor market, with the same caveats as before. In this case, we also notice that couples with a high index 1 are also more educated, likely because educated individuals have children at an older age; their children have perform better in nationally standardized math tests. Regarding index 2, more educated wives are also likely to work and spend time outside home, and the children have higher grades. However, the index is also strongly correlated with children reporting to be less happy and less likely to get along with their mothers. In other words, educated parents' investment in their children's human capital, while fruitful in terms of academic performance, comes at a price, since children appears to resent the corresponding pressure.[16] What is remarkable here is that

[16]This notion had already been mentioned in the literature (Heinrich, 2014; Dinisman and Ben-Arieh, 2016). One explanation is that the parenting style of educated parents is more oriented towards altruism than paternalism: this means that educated parents pressure children to perform well at school, which results in a welfare loss for the child (Doepke and Zilibotti, 2017).

the factors are recovered *exclusively from matching patterns*; neither parental investment nor any output variable are used for their estimation. Yet, the most important factor (beyond age) driving assortativeness appears to be strongly correlated with both children's achievements (positively) and well-being (negatively). This indicates not only that future investments in children's HC are an explicit part of individuals' marital strategies, but also that these aspects are actually crucially important, since the corresponding factor dominates all other patterns (beyond pure age/cohort effects).

As discussed in the previous section, index 3 mainly captures differences in multiple dimensions of health and risk attitude. However, we did not find that children-related outcomes differ in couples with high vs low levels of index 3. On the other hand, we cannot exclude that index 3 is somehow related to other family outcomes that we do not observe and that depend on parents' risk aversion, such as financial decisions on homeownership and portfolio choices.

5 Conclusion

A large literature in both economics, sociology and demography has studied homogamy and measured it using data on marital patterns. In this paper, we present Separable Extreme Value (SEV) models and explain how they compare to other statistical models used to measure homogamy. SEV models are advantageous because i) they are identified with cross-sectional data on one marriage market; ii) they can be used to study multidimensional sorting and can easily handle numerous discrete classes and continuous variables; iii) in the matching function implied by the model, spillovers play a large role, in that an increase in the number of individuals in one class (say, female college graduates) can potentially affect *all* match probabilities.

We show that the SEV approach can generate rich empirical findings by estimating a multidimensional and parametric model borrowed from Dupuy and Galichon (2014) with data from a survey of parents of children attending schools in Campania, a region of Southern Italy. We show that marital patterns are characterized by a high level of homogamy: not only men and women sort on demographic and socioeconomic traits such as age, BMI, height and education, but they also look for partners that share similar health-related behavioral traits and risk attitude. Our estimates are also insightful about the number and nature of the sorting dimensions that can rationalize the marital patterns observed in the data: we find that only a relatively low number of sorting dimensions matter; in other words, one or few indices could well summarize an individual's attractiveness on marriage markets. While the first dimension of sorting mainly captures market

segmentation across age cohorts, the second dimension describes sorting on human capital (HC): educated and health-conscious women tend to marry men with similar traits. Finally, when we look at family outcomes, we find that children of parents with a high level of HC perform better at school, but this comes at a cost, as they are also more likely to report lower levels of subjective well-being and a worse relationship with their mother.

A Computational details

In a sample with N couples, we aim to calculate the unconditional probability of a marriage between a woman of type I and a man of type J - named μ_{IJ} according to our notation - generated by a given surplus function Z^{IJ} and given marginal distributions μ_I^X and μ_J^Y. The logit choice models of the wife and the husband respectively yield:

$$p_{IJ} = \frac{\exp U^{IJ}}{\sum_K \exp U^{IK}} = \frac{\mu_{IJ}}{\mu_I^X}$$
$$p_{JI} = \frac{\exp V^{IJ}}{\sum_K \exp V^{KJ}} = \frac{\mu_{IJ}}{\mu_J^Y}$$

where the last equality on both lines establishes a relationship between the conditional and unconditional probabilities, μ_I^X and μ_J^Y being the marginal distributions of women's and men's types respectively. If every individual belongs to a different class (as is often the case with many discrete categories, or when continuous variables are included in the analysis), then $\mu_I^X = 1/N$ and $\mu_J^Y = 1/N$ for any I and J; if sample weights are available, then μ_I^X is equal to the sample weight of an individual of class I (and so is μ_J^Y).

We proceed from equation (4), which establishes a link between U^{IJ} and V^{IJ}, and write

$$\mu_{IJ} = \exp\left(\frac{Z^{IJ}}{2}\right) a_I^X b_J^Y$$

where a_I^X and b_J^Y are type-specific terms defined as

$$a_I^X = \sqrt{\frac{\mu_I^X}{\sum_K \exp U^{IK}}}$$
$$b_J^Y = \sqrt{\frac{\mu_J^Y}{\sum_K \exp V^{KJ}}}.$$

For a given function Z^{IJ}, we can obtain μ_{IJ} if we can compute a_I^X and b_J^Y for all types. We use the adding-up constraints (3) to set up the following system

$$\mu_I^X = \sum_J \mu_{IJ} = a_I \sum_J \exp\left(\frac{Z^{IJ}}{2}\right) b_J$$
$$\mu_J^Y = \sum_I \mu_{IJ} = b_J \sum_I \exp\left(\frac{Z^{IJ}}{2}\right) a_I$$

whose solution can be obtained through the well-known iterative proportional fitting procedure (IPFP). The properties of the latter ensures that we always (and rapidly) converge to a solution a_I and b_J that is unique up to a scalar; for any Z^{IJ}, there exists a unique μ_{IJ}. Dupuy and Galichon

(2014) provide additional details on this methodology.

B Variable description

In Table 8 we report the survey questions that were used to measure different behavioral traits. In Table 9 we report a complete list of all variables that were used in section 4.3 as "outcome" variables. Note that some outcome variables may never show up in our plots because they are found to be uncorrelated with the estimated indices of attractiveness.

Table 8: Matching variables

Matching variables		
Likes healthy snacks	*At the end of the experiment, we will give you a snack: which one do your prefer?*	1=Nutella sandiwch, 2=Parmesan bar, 3=apple
Smokes	*I smoke*	0=Never, 1=Sometimes, 2=Often
Likes sports	*I do sports*	0=Never, 1=Sometimes, 2=Often
Wears sunscreen	*I wear sunscreen to avoid sunburns*	0=Never, 1=Sometimes, 2=Often
Washes hands	*I wash my hands before eating*	0=Never, 1=Sometimes, 2=Often
Worries about health	*I worry about my health*	0=Never, 1=Sometimes, 2=Often
Wears helmet	*I wear a helmet when riding a moped*	0=Never, 1=Sometimes, 2=Often
Likes safari	*I would go on a safari in the jungle*	0=Never, 1=Sometimes, 2=Often
Fears speed	*I am scared of mopeds riding fast*	0=Never, 1=Sometimes, 2=Often
Likes usual holidays	*I like going on holidays in places I know because it is safer*	0=Never, 1=Sometimes, 2=Often
Likes extreme sports	*I would do extreme sports*	0=Never, 1=Sometimes, 2=Often
Careful when crossing	*I am very careful when crossing the street*	0=Never, 1=Sometimes, 2=Often

Table 9: Outcome variables

Outcome variables	
Number of children	Children report family composition, including number and sex of siblings.
Stays at home (wife)	Inputted: 0=Otherwise, 1=Answers "housewife" when asked about profession.
Hours outside (wife)	Mothers are asked: *If you work, how many hours do you usually spend outside the home?* Possible answers: 0=does not work, 1=3 to 5 hours, 2=6 to 8 hours, 3=more than 8 hours.
Height (child)	Measured by interviewers. We use residuals from regression on child's age.
BMI (child)	Measured by interviewers. We use residuals from regression on child's age.
Year failed	From class register.
Italian grade	From class register; we use deviation from class mean.
Math grade	From class register; we use deviation from class mean.
Italian Invalsi grade	From class register; standardized national test.
Math Invalsi grade	From class register; standardized national test.
Sub. well-being (child)	Children are asked: *How happy are you about your life?* Possible answers are: 1=Very sad, ... 5=Very happy. We use residuals from regression on child's age.
Happy (child)	Children are asked: *Are you happy?* Possible answers are: 0=Never, 1=Sometimes, 2=Often. We use residuals from regression on child's age.
Gets along with mum	Children are asked: *Do you get along with your mum?* Possible answers are: 0=Never, 1=Sometimes, 2=Often
Gets along with dad	Children are asked: *Do you get along with your dad?* Possible answers are: 0=Never, 1=Sometimes, 2=Often

C Additional tables

Table 10: Frequency of children by parents' presence at home

	Father is present		
Mother is present	**0**	**1**	**Total**
0	0	3	3
1	16	613	629
Total	16	616	632

Notes: a parent is present if he/she participates to the survey and/or is reported as living at home by the child.

Table 11: Summary statistics - family outcomes

	mean	sd	count
Number of children	2.1	0.8	632
Wife's hours outside	0.5	0.9	658
Wife stays at home	0.5	0.5	300
Child's height	-0.0	8.9	634
Child's BMI	19.6	3.7	620
Year failed	0.0	0.2	659
Italian grade	-0.0	1.0	617
Math grade	-0.0	1.0	619
INVALSI test grade, Italian	202.3	40.9	184
INVALSI test grade, math	194.3	38.3	179
Child's patience	-0.0	0.5	646
Child's subjective well-being	-0.0	0.7	620
Child is happy	-0.0	0.5	632
Child gets along with mum	1.6	0.5	634
Child gets along with dad	1.7	0.6	637
Child likes healthy snacks	1.7	0.9	642

Notes: some variables are normalized to cleanse them from age effects (e.g., child's subjective well-being). Others are normalized to cleanse them from class effects (e.g., math and Italian non-standardized grades).

Table 12: Rank test for $\hat{A}$ (sample 1)

H_0: $rk(A) = k$	$k = 1$	$k = 2$	$k = 3$
χ^2	70.34	17.34	1.91
df	9	4	1
Rejected?	Yes	Yes	No

Notes: each column reports the statistic resulting from testing the null hypothesis that the rank of $\hat{A}$ is equal to k. We report whether the null hypothesis was rejected at the 5% level. These tests lead us to conclude that sorting occurs on at least 3 orthogonal dimensions.

Table 13: Rank test for $\hat{A}$ (sample 2)

H_0: $rk(A) = k$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$	$k = 11$	$k = 12$	$k = 13$	$k = 14$	$k = 15$
χ^2	280.46	242.59	215.12	179.15	144.42	118.19	93.22	63.73	48.23	31.05	19.50	9.06	4.38	0.71	0.10
df	225	196	169	144	121	100	81	64	49	36	25	16	9	4	1
Rejected?	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No

Notes: each column reports the statistic resulting from testing the null hypothesis that the rank of $\hat{A}$ is equal to k. We report whether the null hypothesis was rejected at the 5% level. These tests lead us to conclude that sorting occurs on at least 5 orthogonal dimensions.

Table 14: Diagonal of estimated affinity matrix (sample 2)

Educ.	Age	Height	BMI	Smokes	Likes sports	Likes healthy snacks	Wears sunscreen	Wash hands	Worries about health	Wears helmet	Likes safari	Fear speed	Likes usual holidays	Extreme sports	Careful when crossing
0.99	**4.37**	**0.30**	**0.57**	**0.51**	**0.29**	**0.71**	**0.51**	**0.60**	**0.35**	**0.68**	**0.58**	**0.53**	**0.53**	0.02	**0.47**
(0.19)	(0.49)	(0.11)	(0.12)	(0.12)	(0.12)	(0.13)	(0.14)	(0.14)	(0.12)	(0.15)	(0.12)	(0.13)	(0.11)	(0.10)	(0.13)

Notes: 196 couples. The table only reports the diagonal elements of the affinity matrix. Standard errors in parentheses: the estimator $\hat{A}$ is asymptotically normally. Bold-faced values are significant at the 5% level.

Table 15: Saliency analysis: men's loading matrix (sample 2)

	Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7	Index 8	Index 9	Index 10	Index 11	Index 12	Index 13	Index 14	Index 15	Index 16
Educ.	**0.21**	**0.45**	0.05	**0.59**	**-0.36**	-0.01	**0.25**	0.04	-0.03	-0.02	-0.02	0.15	0.21	0.28	0.00	-0.25
Age	**0.97**	**-0.09**	-0.01	**-0.14**	**0.07**	-0.02	-0.06	0.02	**-0.07**	-0.00	0.04	**-0.06**	**-0.08**	-0.04	0.04	0.03
Height	**0.06**	**0.11**	-0.02	**0.14**	-0.01	-0.07	**-0.37**	-0.07	**0.58**	0.10	**0.31**	0.08	**0.29**	**-0.48**	-0.17	-0.14
BMI	0.03	0.04	**0.19**	**-0.32**	0.06	**0.11**	**0.50**	**-0.32**	**0.37**	**0.52**	-0.02	0.02	0.16	0.14	0.19	-0.04
Smokes	0.01	-0.06	**-0.14**	**-0.15**	-0.03	**-0.16**	**-0.41**	**-0.46**	**-0.34**	0.14	**-0.25**	0.16	**0.52**	0.21	-0.09	-0.09
Likes sports	0.03	**-0.43**	-0.03	**0.24**	**0.25**	**-0.20**	**-0.12**	**0.17**	**0.38**	0.07	-0.01	**0.40**	-0.10	**0.53**	-0.08	0.06
Likes healthy snacks	-0.01	**0.38**	**0.10**	**-0.14**	**0.12**	**0.14**	**-0.36**	**0.55**	0.01	**0.47**	**-0.23**	-0.21	0.03	0.18	-0.11	0.02
Wears sunscreen	**-0.04**	**0.37**	**0.20**	**-0.38**	**0.12**	**-0.24**	0.01	0.08	**-0.23**	-0.05	**0.46**	**0.54**	-0.14	0.06	-0.07	-0.11
Wash hands	0.02	**-0.32**	**0.12**	**-0.21**	**-0.23**	**0.64**	0.07	**0.24**	-0.03	-0.14	-0.01	**0.27**	0.22	-0.01	-0.34	-0.25
Worries about health	0.01	**-0.21**	-0.03	0.10	0.07	**-0.39**	**0.35**	**0.34**	**-0.22**	**0.29**	**-0.28**	**0.24**	0.21	**-0.47**	-0.11	0.00
Wears helmet	-0.01	**-0.11**	**0.64**	0.03	**-0.19**	-0.07	**-0.28**	-0.04	0.05	-0.04	**-0.34**	0.16	-0.20	-0.16	0.43	-0.24
Likes safari	-0.02	**-0.16**	0.01	**0.39**	0.04	**0.36**	**-0.14**	**-0.17**	**-0.36**	**0.54**	**0.34**	0.15	**-0.23**	-0.13	0.09	0.11
Fear speed	0.03	**0.23**	**-0.39**	0.01	**0.28**	**0.18**	0.02	**-0.22**	0.09	0.01	**-0.41**	**0.23**	**-0.44**	-0.14	-0.15	-0.42
Likes usual holidays	0.01	**0.08**	**0.22**	**0.20**	**0.75**	**0.24**	0.07	0.03	-0.10	**-0.26**	0.00	-0.01	**0.38**	-0.04	0.22	-0.05
Extreme sports	**0.05**	**0.26**	-0.02	-0.00	-0.10	**0.22**	-0.03	-0.07	0.14	-0.12	-0.28	**0.42**	0.04	-0.12	0.05	0.75
Careful when crossing	-0.03	-0.03	**-0.53**	**-0.14**	-0.13	0.09	-0.07	**0.28**	0.03	0.00	0.10	0.17	0.16	-0.01	**0.71**	-0.14
Index share	0.32	0.13	0.10	0.09	0.06	0.06	0.05	0.05	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.00

Notes: the table reports men's singular vectors V and singular values $diag(\Lambda)$ from the singular value decomposition of $\hat{A} = U\Lambda V'$. Bold-faced values are significant at the 5% level; confidence intervals are obtained with 500 bootstrap replications (Milan and Whittaker, 1995). In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

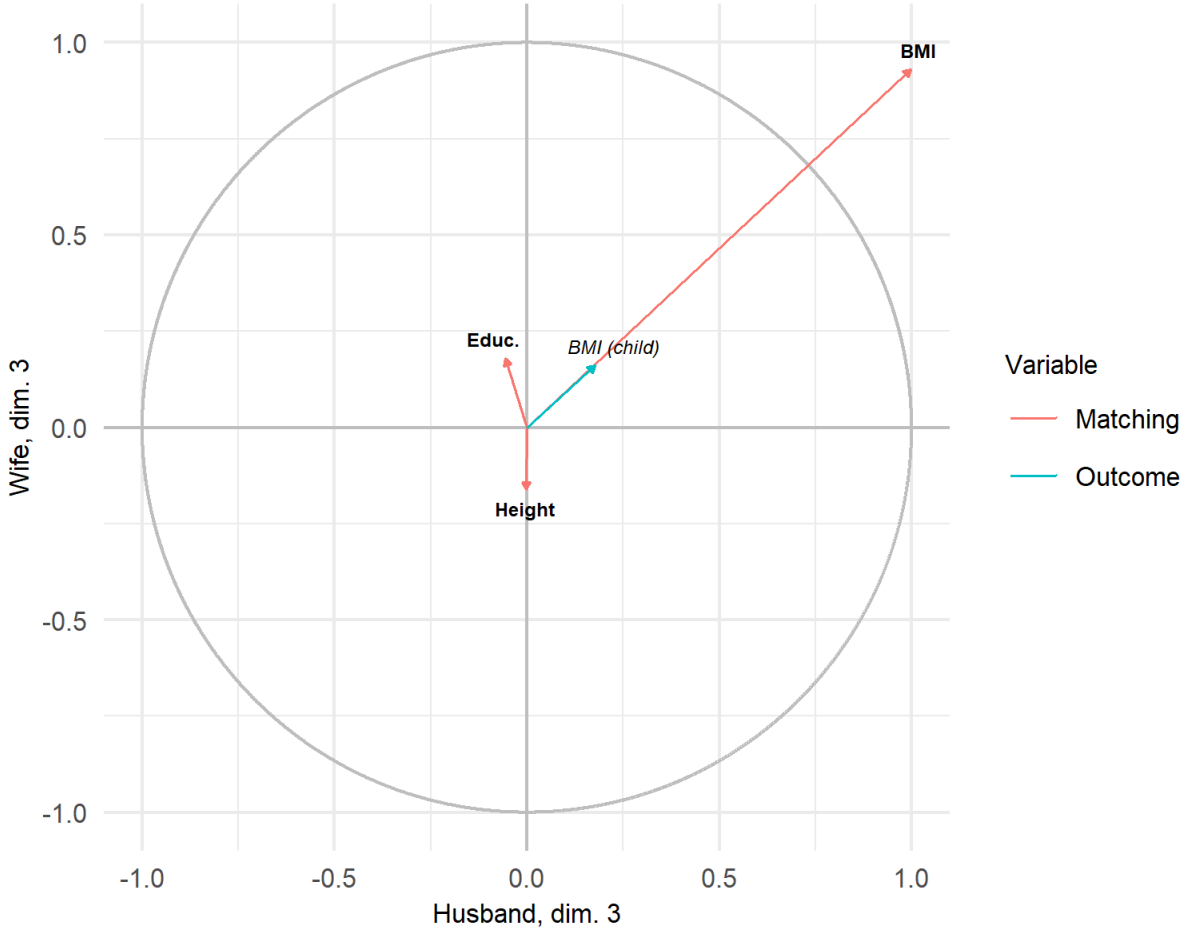
Table 16: Saliency analysis: women's loading matrix (sample 2)

	Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7	Index 8	Index 9	Index 10	Index 11	Index 12	Index 13	Index 14	Index 15	Index 16
Educ.	**0.09**	**0.62**	**0.29**	**0.39**	**-0.21**	**-0.23**	-0.01	-0.09	**0.24**	0.09	0.04	-0.13	**0.28**	0.16	0.25	0.16
Age	**0.97**	**-0.12**	-0.03	**0.06**	**0.08**	**0.08**	-0.05	**0.07**	-0.03	**-0.07**	-0.02	**-0.08**	0.06	-0.02	-0.02	0.04
Height	**0.09**	0.01	-0.07	0.05	**-0.14**	**-0.15**	-0.08	0.02	**0.56**	0.18	0.14	**0.50**	0.04	-0.32	-0.29	-0.36
BMI	**0.06**	-0.01	0.03	**-0.43**	**0.32**	0.09	**0.34**	**-0.27**	**0.30**	**0.38**	-0.00	-0.07	0.25	0.44	-0.04	-0.09
Smokes	0.03	**0.19**	**-0.37**	**-0.18**	-0.09	**-0.14**	**-0.55**	**-0.38**	**-0.19**	-0.02	**-0.43**	0.11	0.18	0.15	-0.00	-0.18
Likes sports	0.02	-0.07	-0.01	**0.23**	0.05	-0.10	-0.06	0.05	**0.19**	0.14	**-0.34**	0.25	**-0.48**	0.41	-0.27	0.46
Likes healthy snacks	0.02	**0.47**	0.04	**-0.13**	**0.15**	**0.18**	**-0.15**	**0.64**	**-0.24**	**0.20**	0.05	0.07	-0.09	0.21	-0.18	-0.29
Wears sunscreen	**0.11**	**0.31**	-0.08	**-0.56**	-0.03	**-0.14**	-0.05	-0.05	0.04	0.14	0.19	0.11	-0.36	-0.31	0.31	0.40
Wash hands	0.01	-0.02	**0.26**	**-0.15**	**-0.51**	**0.73**	-0.08	-0.06	0.08	-0.04	-0.11	**0.24**	0.06	0.08	0.08	0.10
Worries about health	0.01	-0.07	**-0.15**	0.11	0.03	-0.10	**0.34**	**0.17**	**-0.37**	**0.22**	-0.14	**0.57**	**0.42**	-0.07	0.20	0.23
Wears helmet	**0.06**	**-0.24**	**0.57**	-0.11	0.00	**-0.29**	-0.06	0.07	-0.04	0.10	**-0.33**	0.11	-0.22	0.04	0.44	-0.37
Likes safari	-0.02	**-0.13**	**0.11**	**0.32**	**0.23**	**0.22**	**-0.37**	**-0.29**	**-0.22**	**0.62**	**0.31**	0.00	-0.08	-0.12	0.07	0.03
Fear speed	-0.01	**0.20**	**-0.40**	**0.28**	**0.24**	**0.36**	**0.24**	-0.03	**0.21**	0.01	**-0.32**	-0.03	**-0.27**	-0.16	0.43	-0.23
Likes usual holidays	**-0.11**	-0.02	**0.20**	-0.04	**0.60**	**0.13**	**-0.39**	**0.14**	**0.28**	**-0.36**	-0.03	**0.19**	**0.28**	-0.08	0.13	0.24
Extreme sports	**0.07**	0.05	-0.07	0.09	0.05	-0.02	0.05	**-0.23**	-0.13	**-0.37**	**0.52**	**0.41**	-0.22	0.45	0.21	-0.20
Careful when crossing	**-0.05**	**-0.35**	**-0.38**	-0.01	**-0.24**	-0.08	**-0.27**	**0.41**	**0.27**	0.16	0.20	-0.16	0.13	0.28	0.41	0.05
Index share	0.32	0.13	0.10	0.09	0.06	0.06	0.05	0.05	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.00

Notes: the table reports women's singular vectors U and singular values $diag(\Lambda)$ from the singular value decomposition of $\hat{A} = U\Lambda V'$. Bold-faced values are significant at the 5% level; confidence intervals are obtained with 500 bootstrap replications (Milan and Whittaker, 1995). In the last line, each value of $diag(\Lambda)$ can be interpreted as the relative importance of each sorting dimension.

D Additional figures

Figure 6: Correlation between index 3, matching variables and outcomes (sample 1)



Notes: we plot correlation rates of both matching variables and outcome variables with the husband's third matching index (x-axis) and the wife's (y-axis). See notes to figure 1.

References

Gary S Becker. A theory of marriage: Part i. Journal of Political economy, 81(4):813–846, 1973.

Debra L Blackwell and Daniel T Lichter. Homogamy among dating, cohabiting, and married couples. The Sociological Quarterly, 45(4):719–737, 2004.

Ann-Renée Blais and Elke U Weber. A domain-specific risk-taking (dospert) scale for adult populations. Judgment and Decision making, 1(1), 2006.

Milan Bouchet-Valat. Changes in educational, social class and social class of origin homogamy in france (1969-2011). Revue française de sociologie, 55(3):459–505, 2014.

Pierre-André Chiappori, Bernard Salanié, and Yoram Weiss. Partner choice, investment in children, and the marital college premium. American Economic Review, 107(8):2109–67, 2017.

Pierre-André Chiappori, Monica Costa Dias, and Costas Meghir. Changes in assortative matching: Theory and evidence for the us. Technical report, National Bureau of Economic Research, 2020.

Eugene Choo and Aloysius Siow. Who marries whom and why. Journal of political Economy, 114 (1):175–201, 2006.

Edoardo Ciscato, Alfred Galichon, and Marion Goussé. Like attract like? a structural comparison of homogamy across same-sex and different-sex households. Journal of Political Economy, 128 (2):740–781, 2020.

Daniela Del Boca, Christopher Flinn, and Matthew Wiswall. Household choices and child development. Review of Economic Studies, 81(1):137–185, 2014.

Tamar Dinisman and Asher Ben-Arieh. The characteristics of children's subjective well-being. Social indicators research, 126(2):555–569, 2016.

Matthias Doepke and Fabrizio Zilibotti. Parenting with style: Altruism and paternalism in intergenerational preference transmission. Econometrica, 85(5):1331–1371, 2017.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. The intergenerational transmission of risk and trust attitudes. The Review of Economic Studies, 79(2):645–677, 2012.

Arnaud Dupuy and Alfred Galichon. Personality traits and the marriage market. Journal of Political Economy, 122(6):1271–1319, 2014.

Arnaud Dupuy, Alfred Galichon, and Yifei Sun. Estimating matching affinity matrices under low-rank constraints. Information and Inference: A Journal of the IMA, 8(4):677–689, 2019.

Eurostat. Education and training eurostat, european community household panel (echp), gender equality indicators, 2016. URL https://ec.europa.eu/eurostat/web/microdata/european-community-household-panel.

Raquel Fernández and Richard Rogerson. Sorting and long-run inequality. Quarterly Journal of Economics, 116(4):1305–1341, 2001.

Raquel Fernández, Nezih Guner, and John Knowles. Love and money: A theoretical and empirical analysis of household sorting and inequality. Quarterly Journal of Economics, 120(1):273–344, 2005.

Pilar Gonalons-Pons and Christine R Schwartz. Trends in economic homogamy: changes in assortative mating or the division of labor in marriage? Demography, 54(3):985–1005, 2017.

Jeremy Greenwood, Nezih Guner, and John A Knowles. More on marriage, fertility, and the distribution of income. International Economic Review, 44(3):827–862, 2003.

Jeremy Greenwood, Nezih Guner, Georgi Kocharkov, and Cezar Santos. Marry your like: Assortative mating and income inequality. American Economic Review, 104(5):348–53, 2014.

Carla Guerriero, John Cairns, Fabrizio Bianchi, and Liliana Cori. Are children rational decision makers when they are asked to value their own health? a contingent valuation study conducted with children and their parents. Health Economics, 27(2):e55–e68, 2018.

Carolyn J Heinrich. Parents' employment and children's wellbeing. The future of children, pages 121–146, 2014.

Luis Milan and Joe Whittaker. Application of the parametric bootstrap to models that incorporate a singular value decomposition. Journal of the Royal Statistical Society: Series C (Applied Statistics), 44(1):31–49, 1995.

Zhenchao Qian and Samuel H Preston. Changes in american marriage, 1972 to 1987: Availability and forces of attraction by age and education. American Sociological Review, 58(4):482–495, 1993.

Robert Schoen. The harmonic mean as the basis of a realistic two-sex marriage model. Demography, 18(2):201–216, 1981.

Christine R Schwartz and Hongyun Han. The reversal of the gender gap in education and trends in marital dissolution. American Sociological Review, 79(4):605–629, 2014.

Christine R Schwartz and Robert D Mare. Trends in educational assortative marriage from 1940 to 2003. Demography, 42(4):621–646, 2005.

Elke U Weber, Ann-Renee Blais, and Nancy E Betz. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. Journal of Behavioral Decision Making, 15(4):263–290, 2002.